

Kurzantwortfragen oder Multiple-choice-Fragen?

Allgemeinmedizinische Facharztprüfung 1. Teil

P. Schläppi, D. Hofer, R. Hofer, R. Bloch

Zusammenfassung

Hintergrund und Fragestellung

Die Schweizerische Gesellschaft für Allgemeinmedizin SGAM verwendet Kurzantwortfragen (KAF), um das aktiv formulierbare Wissen der Facharztprüflinge zu beurteilen. Die SGAM findet dieses Wissen praxisrelevanter als passiv wiedererkanntes, wie es beispielsweise mit Multiple-choice-Fragen (MC) überprüft werden kann. Es stellte sich aber die Frage, ob die Leistungen der Prüfungsteilnehmerinnen und -teilnehmer mit MC-Fragen wirklich anders beurteilt würden als mit KAF.

Methode

Einstündige Testprüfung mit fünf Prüfungsfällen im Anschluss an die zweistündige Facharztprüfung (KAF) 1998. Leistungsvergleich zweier prospektiv parallelisierter Gruppen aus den 107 deutschsprachigen Prüfungskandidatinnen und -kandidaten 1998. Eine Gruppe machte die Prüfung in der KAF-Version, die andere in der MC-Version (zufällige Gruppenzuteilung). Die beiden Versionen wurden auf Fallenebene in einem konsensuellen Äquivalenzprozess vergleichbar gemacht. Die Leistungen der Kandidatinnen und Kandidaten wurden in den sich möglichst genau korrespondierenden Antwortitems der beiden Versionen verglichen.

Resultate

Signifikanter Schwierigkeitsunterschied der Äquivalenzitems: MC-Version im Mittel 69% richtige Antworten, in der KAF-Version 51% (Mann-Whitney U, $p < 0.01$).

Schlussfolgerung

Würde vergleichbares Wissen in der schweizerischen Facharztprüfung Allgemeinmedizin statt mittels KAF mit MC-Fragen geprüft, wäre die Prüfung für die Kandidatinnen und Kandidaten höchstwahrschein-

lich deutlich leichter zu lösen. Das effektiv in der Praxis anwendbare Wissen der Kandidatinnen und Kandidaten würde mit MC um etwa 20% überschätzt und möglicherweise auch falsch eingestuft.

Einleitung

Die Schweizerische Gesellschaft für Allgemeinmedizin SGAM beurteilt das allgemeinärztliche Wissen in der Facharztprüfung mit Kurzantwortfragen (KAF), weil damit aktives Wiedergeben, Anwenden und Interpretieren von Kenntnissen praxisrelevanter geprüft werden können als z.B. mit der MC-Methode. Bei dieser Methode steht ja das Wiedererkennen von meist partialisiertem Faktenwissen im Vordergrund.

Weil der Aufwand für die Neueinführung dieser Facharztprüfung aber beträchtlich ist, stellte sich u.a. die Frage, ob nicht doch auf eine Prüfung mit bewährter MC-Methodik umgestellt werden sollte, v.a. dann natürlich, wenn damit die Leistungen der Prüfungskandidatinnen und -kandidaten etwa gleich beurteilt würden.

Aus der Literatur gibt es sowohl Hinweise darauf, dass es einen Unterschied macht, ob man einen gleichen Wissensinhalt mit MC oder KAF prüft [1, 2], als auch Angaben, die keine Evidenz für einen Messunterschied ausmachen konnten [3, 4].

Fragestellung

Bei den Kandidatinnen und Kandidaten der Facharztprüfung Allgemeinmedizin 1998 sollte mit einer Testprüfung die Frage beantwortet werden, ob sich ein Unterschied in der Leistungsbeurteilung zeigt, wenn identische Wissensinhalte mit KAF oder mit MC-Fragen geprüft werden.

Methode

Die Prüfungsverordnung für die allgemeinmedizinischen Facharztprüfungen sieht nur KAF als schriftliche Prüfungsmethode vor. Es musste also eine Testphase zur normalen Prüfung angefügt werden, um auch MC-Fragen verwenden zu können. Die Testdauer war aus logistischen Gründen auf eine Stunde beschränkt. Das war zu kurz, dass zwei Gruppen eine Prüfung hätten absolvieren können, die aus MC- und KA-Fragen bestanden hätte. Die Kandidatinnen und Kandidaten konnten auch nicht in der gleichen Sitzung beide identischen Prüfungsformen durchlaufen, da ja MC-Wahlmöglichkeiten die KAF-Antworten implizieren (oder umgekehrt).

Somit war klar, dass diese Testprüfung grundsätzlich nicht zuverlässige und gültige Messeigenschaften erreichen konnte. Hingegen sollte die Fragestellung doch hinreichend beantwortet werden können, wenn nach einem Äquivalenzprozess Fälle und Fragen direkt verglichen werden und schliesslich das Resultat deutlich ausfällt.

Korrespondenz:

Dr. med. Peter Schläppi

Institut für Aus-, Weiter- und Fortbildung IAWF

Medizinische Fakultät der Universität Bern

Inselspital 37a

CH-3010 Bern

E-mail peter.schlaeppi@iae.unibe.ch

Im Fallbeispiel (Abb. 1) konnten direkt verglichen werden:

Tabelle 1

Antworten Version KAF	Antworten Version MC
Stuhlbakteriologie (Salmonellen, Shigellen)	1. D bakteriologische Stuhluntersuchung auf Salmonellen und Shigellen
Erweiterte Anamnese: Antibiotikatherapie? Blut im Stuhl? Psychische Störung?	2. D. Antibiotikatherapie 2. B. Frage nach Blut im Stuhl 4. C. psychische Störung
Abdominaluntersuchung	3. A. Abdominaluntersuchung
bakteriell, Protozoen, Parasiten	4. A. Protozoeninfekt
Medikamente	2. A. Medikamentengebrauch
Nahrungsmittelallergie	2. C. Nahrungsmittelallergien
Stoffwechselerkrankung (Hyperthyreose, Pankreasleiden)	5. A. Hypothyreose [D. Pankreatitis]
Lebensmittelintoxikation	5. [B. Lebensmittelintoxikation]
Chronische Entzündung (M. Crohn)	5. [C. M. Crohn]
Laxantienabusus	5. [E. Laxantienabusus]

Im Anschluss an die zweistündige Facharztprüfung 1998 absolvierte schliesslich die eine Hälfte der Prüflinge eine MC-Testprüfung, die andere eine Testprüfung mit KAF. Die angemeldeten 107 deutschsprachigen Kandidatinnen und Kandidaten wurden prospektiv in zwei Gruppen (MC 54, KAF 53) eingeteilt, «gematched» nach Alter, Geschlecht, Diplomjahr, Anzahl Jahre in Weiterbildung und nach der Anzahl Monate, die die Kandidatinnen und Kandidaten in einer Praxis absolviert hatten. Per Los wurde der einen Gruppe die MC-Prüfung, der andern die KAF-Prüfung zugeteilt. Die Kandidatinnen und Kandidaten waren über Ziel und Zweck der Untersuchung im voraus informiert.

Die Hauptschwierigkeit der Untersuchung bestand darin, beiden Gruppen möglichst identische Prüfungsaufgaben vorzulegen. Es ist kaum möglich, einen Vergleichstest zu konstruieren, bei dem eine Wahlantwort inhaltlich genau einer Kurzantwort entspricht (und umgekehrt).

Bei den MC-Fragen verwendeten wir zwei unterschiedliche Typen (vgl. rechte Spalte in Abb. 1). Beim einen (Typ A) werden zu einer Fragestellung fünf Antworten vorgegeben, von welchen nur eine auszuwählen und zu bezeichnen ist, die einzig richtige oder die am meisten zutreffende. Beim andern (Typ K') werden vier Aussagen vorgelegt, deren Richtigkeit einzeln beurteilt werden muss. Da in KAF meist mehrere richtige Antworten gefordert werden, eignete sich der Typ K' zur «Übersetzung» in die MC-Form eher besser, weil hier ebenfalls in einer Frage mehrere richtige Aussagen angeboten werden können. Alternativ wurde versucht, das in einer KAF geprüfte Wissen mit mehreren Typ-A-Fragen abzudecken.

Gewisse methodische Schwierigkeiten sind nicht lösbar. Zum Beispiel benötigt in den MC-Typen auch das Erkennen, welche Aussagen falsch sind, Wissen, das in den KAF nicht überprüft wird. Oder aber die Tatsache, dass Raten in KAF nicht möglich ist.

Für diese Untersuchung wurden in einem Äquivalenz- und Konsensusverfahren zuerst fünf bereits verwendete KAF-Fälle herangezogen (aus Prüfung 97). Diese Fälle hatten das offizielle Revisionsverfahren der Facharztprüfung durchlaufen (Prozess mit ca. 20 Fachärztinnen und -ärzten inkl. Prüfungskommission). Es bestand ein definierter Schlüssel mit richtigen Antworten. Anschliessend wurden alle essentiellen und möglichst viele weitere richtige Antworten dieses Schlüssels in MC-Fragen des Typs A und K' gefasst. Bei den «Ablenkern» der Typ-A-Fragen wurden möglichst keine neuen Wissensinhalte verwendet, sondern richtige Antworten zu KAF so angefügt, dass sie weniger wahrscheinlich sind. Drei Experten und eine Expertin aus dem Facharztprüfungspanel überprüften Inhalt und Identität der so verfassten neuen MC-Fragen. Mitarbeiter des IAWF bereinigten die formale Qualität.

Am nachfolgenden Beispiel (Abb. 1), einem der fünf Fälle aus der Untersuchung, ist das Resultat dieses Äquivalenzverfahrens aufgezeigt. In der KAF-Version sind zu jeder Frage die Antworten des Schlüssels eingefügt.

In dieser Art lagen schliesslich fünf Fälle in beiden Methoden vor, bei denen vergleichbare Wissensitems erfragt wurden: Den 22 Wissensitems der MC-Version (= 22 MC-Fragen, 15 Typ K' und 7 Typ A) standen dabei 40 zugeordnete KAF-Antwortmöglichkeiten gegenüber.

Abbildung 1

Frau Blickensdorfer, eine bis anhin gesunde Angestellte im Gastgewerbe, sucht Sie in Ihrer Praxis auf und klagt über eine neu aufgetretene, seit zwei Wochen bestehende Diarrhøe.

Version KAF (essentielle Antworten *kursiv* gedruckt)

- Welche diagnostischen Massnahmen schlagen Sie bei Frau Blickensdorfer vor? (mindestens 3).
 - Stuhlbakteriologie (Salmonellen, Shigellen)*
 - Erweitere Anamnese: Ferienreise? Antibiotikatherapie? Blut im Stuhl?
 - Labor: Hb, Lc, BSR/CRP, okkultes Blut
 - Abdominaluntersuchung
- Nennen Sie mindestens 6 häufige Ursachen für eine Diarrhøe (unabhängig von der Situation von Frau Blickensdorfer).
 - bakteriell, Protozoen, Parasiten*
 - viral*
 - Medikamente, Laxantienabusus*
 - psychisch (Colon irritabile)*
 - Nahrungsmittelallergie
 - Lebensmittelintoxikation
 - Chronische Entzündung (M. Crohn, Colitis ulcerosa)
 - Stoffwechselerkrankung (Hyperthyreose, Pankreasleiden)
- Nennen Sie zu 4 von Ihnen unter 2. erwähnten Diarrhøe-ursachen mindestens eine diagnostische oder therapeutische Massnahme.
 - evtl. Stuhlbakteriologie, Antibiotikum
 - Wasser- und Elektrolytersatz, evtl. Loperamid
 - Medikamentenstop
 - Gesprächstherapie
 - Diät
 - evtl. weitere Abklärung durch Kantonsarzt
 - Endoskopie oder Dünndarmpassage mit Kontrast, evtl. Prednison, Asacol
 - Labor: mindestens TSH, BZ, Amylase

Version MC (richtige Antworten fettgedruckt)

- Welches ist die *wichtigste* Massnahme, die Sie jetzt der Patientin vorschlagen?

Typ A

 - Stuhluntersuchung auf Parasiten, Hb, Leukozytenzahl, Blutsenkung**
 - Hb, Leukozytenzahl, serologische Untersuchungen auf Salmonellen und Shigellen
 - Endoskopie
 - bakteriologische Stuhluntersuchung auf Salmonellen und Shigellen**
 - Hospitalisation
- Worüber sprechen Sie mit Frau Blickensdorfer in dieser Konsultation?

Typ K'

 - Medikamentengebrauch**
 - Frage nach Blut im Stuhl**
 - Nahrungsmittelallergien**
 - Colitis ulcerosa
- Zu den weiteren Massnahmen im Rahmen dieser Konsultation gehört:

Typ K'

 - Abdominaluntersuchung**
 - Verordnung einer salzarmen Diät
 - Meldung an den Kantonsarzt
 - Anmeldung zur Magen-Darm-Passage
- Zu den möglichen Ursachen einer Diarrhøe zählt:

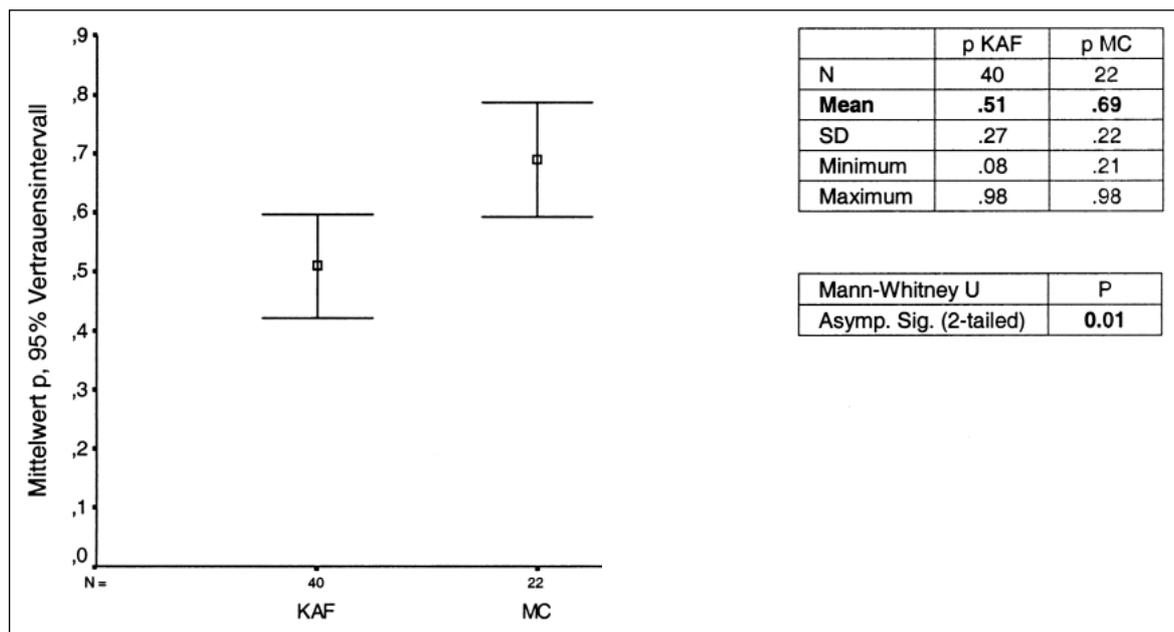
Typ K'

 - Protozoeninfekt**
 - Opiatmedikation
 - psychische Störung**
 - Antibiotikatherapie**
- Die folgenden Krankheitszustände gehören ebenfalls zu den möglichen Ursachen einer Diarrhøe, *ausser*:

Typ A

 - Hypothyreose**
 - Lebensmittelintoxikation
 - M. Crohn
 - Pankreatitis
 - Laxantienabusus

Abbildung 2



Resultate

Drei Kandidatinnen und Kandidaten traten nicht zur Facharztprüfung an. Mit Ausnahme eines Kandidaten lösten sämtliche anwesenden Kandidatinnen und Kandidaten die Testprüfung (Dropout je 2 Kandidatinnen und Kandidaten in beiden Gruppen, resp. 3,7%). Es konnten schliesslich die Resultate von 52 Kandidatinnen und Kandidaten der MC-Gruppe und 51 der KAF-Gruppe ausgewertet werden.

Die Testprüfung wurde am IAWF ausgewertet, wobei die Korrektur der KAF nach dem bestehenden Antwortschlüssel von einem Facharzt Allgemeinmedizin durchgeführt wurde.

Der Vergleich der im Mittel richtig beantworteten MC-Fragen (mittlere Schwierigkeit p MC) mit den entsprechenden KAF-Antworten (mittlere Nennhäufigkeit p KAF) ergab folgendes Bild:

Die vergleichbaren Wissensitems waren also für die Kandidatinnen und Kandidaten in der KAF-Version signifikant schwieriger zu beantworten als in der MC-Version.

Diskussion

Der deutliche und auch signifikante Unterschied bei den Äquivalenzitems lässt doch mit grosser Wahrscheinlichkeit darauf schliessen, dass auch bei einer repräsentativen Fallzahl und angemessener Testdauer ein Unterschied bestehen würde, ob vergleichbare

Kenntnisprüfungsinhalte mit der MC-Methode oder mit KAF geprüft werden. Der Entwicklungs- und Konsensprozess zur Äquivalenz der beiden Versionen scheint sowohl inhaltlich wie formal hinreichend, dass nicht unterschiedliche Wissensinhalte oder die formale Qualität die Differenz erklären können. Die Differenz betrug fast 20%. Etwa um diesen Prozentsatz müsste man also die Kandidatinnen- und Kandidatenleistungen korrigieren, wenn die Prüfung nur aus MC-Fragen bestehen würde und die aktive Wissenskompetenz der Allgemeinmedizinerinnen und -mediziner daraus abzuleiten wäre.

Ist dieser deutliche Unterschied aber alleine durch die unterschiedlichen kognitiven Anforderungen zu erklären oder spielen nicht noch andere Faktoren eine Rolle? In anderen Bereichen (Mathematik [3], Computerwissenschaften [4]) scheint der gedächtnistheoretische Unterschied zwischen den Methoden nicht von Bedeutung zu sein.

Eine mögliche Erklärung ist, dass sich die MC-gewohnten Medizinstudierenden ihr Wissen im Studium hauptsächlich in der passiven Art des Wiedererkennens aneignen, weil es schliesslich in der Ausbildung meist auch so – d.h. mit MC-Fragen – geprüft wird. Dass die Methodik wichtiger Studienprüfungen das Lernverhalten steuert, ist bekannt [5]. Solange also «zählende» Medizinalprüfungen hauptsächlich aus MC-Fragen bestehen, werden Studierende ihr Lernverhalten nicht ändern und können auch praxisrelevantes Wissen zuwenig aktiv formulieren.

Möglicherweise ist aber auch das heutige Studiensystem an sich mitverantwortlich dafür. Angehende Ärztinnen und Ärzte sind oft zuwenig aktiv in den Lernprozess einbezogen, hören mehrheitlich passiv den Dozierenden in Vorlesungen zu oder schauen praktisch tätigen Ärztinnen und Ärzten nur «über die Schulter». Aktiv herausgeforderte Studierende könnten wahrscheinlich auch an einer Facharztprüfung mit KAF ihr Wissen besser aktiv formulieren. Die gegenwärtige Neuordnung des Studiums nach dem Prinzip des problemorientierten Lernens bringt hier möglicherweise künftig eine Änderung, die sich in einigen Jahren auch in einer Facharztprüfung mit KAF niederschlagen könnte.

Wichtig wäre im übrigen zu untersuchen, ob es je nach Methode unterschiedliche Rangordnungen bei den Kandidatinnen- und Kandidatenleistungen gäbe. Mit andern Worten: würde die Facharztprüfung nur mit MC-Fragen gemacht, fielen möglicherweise «falsche» Kandidatinnen und Kandidaten durch, wenn das aktiv formulierbare Wissen als relevant angesehen wird. Es ist auch zu vermuten, dass in den beiden Methoden die Rückmeldung unterschiedlich ausfallen würde, was die Kandidatinnen und Kandidaten in einzelnen Bereichen wissen und was weniger. Für solche Fragestellungen müsste aber eine gemischte KAF/MC-Prüfung für alle Kandidatinnen und Kandidaten in ausreichender Fallzahl und Testdauer durchgeführt werden.

Schlussfolgerungen

Vergleichbare Prüfungsaufgaben beantworten allgemeinmedizinische Facharztprüfungskandidatinnen und -kandidaten in der MC-Form besser als mit KAF-Form. Der Unterschied ist ausgeprägt und wirft die Frage auf, ob heute die Medizinstudierenden im Studium hauptsächlich nach dem Lernprinzip des Wiedererkennens von Wissen vorgehen und praxisrelevantes Wissen zu wenig aktiv formulieren lernen, wie sie es als Praktizierende brauchen werden.

Literatur

- 1 Newble DI, Baxter A, Elmslie RG. A comparison of multiple-choice tests and free-response tests in examination of clinical competence. *Med Educ* 1979;13:263-8.
- 2 Rabinowitz HK, Hojat M. A comparison of the Modified Essay Question and the Multiple Choice Question Formats: Their relationship to clinical performance. *Fam Med* 1989; 5:364-7.
- 3 Schwarz B. Zur Qualität inhaltlich äquivalenter Kurzantwort- und Multiple Choice-Items und Testformen. *Psychol Erz Unterr* 1983;30:200-9.
- 4 Bennett RE, Rock DA, Wang M. Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement* 1991;28 (1):77-92.
- 5 Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1993;17:165-71.